**Overview**

When a quarterback takes a snap and drops back to pass, what happens next may seem like chaos. As offensive players move in various patterns, the defense works together to prevent successful pass completions and then to quickly tackle receivers that do catch the ball. In this year’s Kaggle competition, your goal is to use data science to better understand the schemes and players that make for a successful defense against passing plays.



In American football, there are a plethora of defensive strategies and outcomes. The National Football League (NFL) has used previous Kaggle competitions to focus on offensive plays, but as the old proverb goes, “defense wins championships.” Though metrics for analyzing quarterbacks, running backs, and wide receivers are consistently a part of public discourse, techniques for analyzing the defensive part of the game trail and lag behind. Identifying player, team, or strategic advantages on the defensive side of the ball would be a significant breakthrough for the game.

This competition uses NFL’s Next Gen Stats data, which includes the position and speed of every player on the field during each play. You’ll employ player tracking data for all drop-back pass plays from the 2018 regular season. The goal of submissions is to identify unique and impactful approaches to measure defensive performance on these plays. There are several different directions for participants to ‘tackle’ (ha)—which may require levels of football savvy, data aptitude, and creativity. As examples:

* What are coverage schemes (man, zone, etc) that the defense employs? What coverage options tend to be better performing?
* Which players are the best at closely tracking receivers as they try to get open?
* Which players are the best at closing on receivers when the ball is in the air?
* Which players are the best at defending pass plays when the ball arrives?
* Is there any way to use player tracking data to predict whether or not certain penalties – for example, defensive pass interference – will be called?
* Who are the NFL’s best players against the pass?
* How does a defense react to certain types of offensive plays?
* Is there anything about a player – for example, their height, weight, experience, speed, or position – that can be used to predict their performance on defense?

What does data tell us about defending the pass play? You are about to find out.

**Data Description**

The 2021 Big Data Bowl data contains player tracking, play, game, and player level information for all possible passing plays during the 2018 regular season. For purposes of this event, passing plays are considered to be ones on a pass was thrown, the quarterback was sacked, or any one of five different penalties was called (defensive pass interference, offensive pass interference, defensive holding, illegal contact, or roughing the passer). On each play, linemen (both offensive and defensive) data are not provided. The focus of this year's contest is on pass coverage.

Here, you'll find a summary of each data set in the 2021 Data Bowl, a list of *key* variables to join on, and a description of each variable.

File descriptions

Game data: The games.csv contains the teams playing in each game. The *key* variable is gameId.

Player data: The players.csv file contains player-level information from players that participated in any of the tracking data files. The *key* variable is nflId.

Play data: The plays.csv file contains play-level information from each game. The *key* variables are gameId and playId.

Tracking data: Files week[week].csv contain player tracking data from all games in week [week]. The *key* variables are gameId, playId, and nflId. There are 17 weeks to a typical NFL Regular Season, and thus 17 data frames with player tracking data are provided.

Game data

* gameId: Game identifier, unique (numeric)
* gameDate: Game Date (time, mm/dd/yyyy)
* gameTimeEastern: Start time of game (time, HH:MM:SS, EST)
* homeTeamAbbr: Home team three-letter code (text)
* visitorTeamAbbr: Visiting team three-letter code (text)
* week: Week of game (numeric)

Player data

* nflId: Player identification number, unique across players (numeric)
* height: Player height (text)
* weight: Player weight (numeric)
* birthDate: Date of birth (YYYY-MM-DD)
* collegeName: Player college (text)
* position: Player position (text)
* displayName: Player name (text)

Play data

* gameId: Game identifier, unique (numeric)
* playId: Play identifier, not unique across games (numeric)
* playDescription: Description of play (text)
* quarter: Game quarter (numeric)
* down: Down (numeric)
* yardsToGo: Distance needed for a first down (numeric)
* possessionTeam: Team on offense (text)
* playType: Outcome of dropback: sack or pass (text)
* yardlineSide: 3-letter team code corresponding to line-of-scrimmage (text)
* yardlineNumber: Yard line at line-of-scrimmage (numeric)
* offenseFormation: Formation used by possession team (text)
* personnelO: Personnel used by offensive team (text)
* defendersInTheBox: Number of defenders in close proximity to line-of-scrimmage (numeric)
* numberOfPassRushers: Number of pass rushers (numeric)
* personnelD: Personnel used by defensive team (text)
* typeDropback: Dropback categorization of quarterback (text)
* preSnapHomeScore: Home score prior to the play (numeric)
* preSnapVisitorScore: Visiting team score prior to the play (numeric)
* gameClock: Time on clock of play (MM:SS)
* absoluteYardlineNumber: Distance from end zone for possession team (numeric)
* penaltyCodes: NFL categorization of the penalties that ocurred on the play. For purposes of this contest, the most important penalties are Defensive Pass Interference (DPI), Offensive Pass Interference (OPI), Illegal Contact (ICT), and Defensive Holding (DH). Multiple penalties on a play are separated by a ; (text)
* penaltyJerseyNumber: Jersey number and team code of the player commiting each penalty. Multiple penalties on a play are separated by a ; (text)
* passResult: Outcome of the passing play (C: Complete pass, I: Incomplete pass, S: Quarterback sack, IN: Intercepted pass, text)
* offensePlayResult: Yards gained by the offense, excluding penalty yardage (numeric)
* playResult: Net yards gained by the offense, including penalty yardage (numeric)
* epa: Expected points added on the play, relative to the offensive team. Expected points is a metric that estimates the average of every next scoring outcome given the play's down, distance, yardline, and time remaining (numeric)
* isDefensivePI: An indicator variable for whether or not a DPI penalty ocurred on a given play (TRUE/FALSE)

Tracking data

Each of the 17 week[week].csv files contain player tracking data from all passing plays during Week [week] of the 2018 regular season. Nearly all plays from each [gameId] are included; certain plays or games with insufficient data are dropped. Each team and player plays no more than 1 game in a given week.

* time: Time stamp of play (time, yyyy-mm-dd, hh:mm:ss)
* x: Player position along the long axis of the field, 0 - 120 yards. See Figure 1 below. (numeric)
* y: Player position along the short axis of the field, 0 - 53.3 yards. See Figure 1 below. (numeric)
* s: Speed in yards/second (numeric)
* a: Acceleration in yards/second^2 (numeric)
* dis: Distance traveled from prior time point, in yards (numeric)
* o: Player orientation (deg), 0 - 360 degrees (numeric)
* dir: Angle of player motion (deg), 0 - 360 degrees (numeric)
* event: Tagged play details, including moment of ball snap, pass release, pass catch, tackle, etc (text)
* nflId: Player identification number, unique across players (numeric)
* displayName: Player name (text)
* jerseyNumber: Jersey number of player (numeric)
* position: Player position group (text)
* team: Team (away or home) of corresponding player (text)
* frameId: Frame identifier for each play, starting at 1 (numeric)
* gameId: Game identifier, unique (numeric)
* playId: Play identifier, not unique across games (numeric)
* playDirection: Direction that the offense is moving (text, left or right)
* route: Route ran by offensive player (text)

